

Statistical approaches to the evaluation of the impact  
of vaccination programs: a case study exploring  
rotavirus vaccination in New South Wales

Nicole Mealing

A thesis submitted in fulfilment of the requirements for the

Degree of Doctor of Philosophy

Faculty of Health

The University of Technology Sydney

2018



## **Certificate of Original Authorship**

I, Nicole Mealing, declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Faculty of Health at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by an Australian Government Research Training Program Scholarship.

Production Note:

Signature of Student: Signature removed prior to publication.

Date: 22/03/2018

## Acknowledgements

First and foremost I'd like to thank my supervisors, Professor Andrew Hayen and Associate Professor Anthony Newall. It is a proud moment to see this thesis come together, so thank you for your expertise, direction and feedback over my candidature. Special mention must also go to Dr Robin Turner, not just a good friend but a knowledgeable and supportive biostatistician who offered her skills whenever needed. Thank you especially for your input over this past year, I shall miss our café sessions.

Obtaining data for this thesis proved difficult and time consuming with multiple potential data sources approached. Sincere thanks to Sarah Thackway and her team in the Centre for Epidemiology and Evidence at the NSW Ministry of Health who provided the NSW emergency department and hospital data used in this thesis. These data came from the Emergency Department Data Collection and the Admitted Patient Data Collection maintained by the NSW Ministry of Health. Many thanks to Terrence Flood and Glenys Conner from the Pathology West network who provided the rotavirus pathology data used in this thesis.

This research began at the University of New South Wales and was completed at the University Technology Sydney. It was supported by an Australian Government Research Training Program Scholarship, formerly known as an Australian Postgraduate Award. I was fortunate to attend local and international conferences along with a mathematical modelling course with support from the University of New South Wales Australia Postgraduate Research Support Scheme, University of Technology Sydney's Health Services and Practice Research Development Award and as part of my involvement in an Australian Research Council linkage project grant (LP120200043).

There were many friends who were part of this journey in big and small ways. To Therese, thank you for all the support from start to finish and for editing this thesis. To Cameron, cheers for your advice on extracting the laboratory data into a workable dataset. To Ellen, you were right. Thank you for your friendship, support and the many adventures, even though I didn't listen to you. To Saritha, your patience is incredible. This thesis would not have been completed 'so soon', or possibly even at all, if it wasn't for you. To Mum and Vanessa, for their encouragement and support along the way, especially during an unexpected move. To Bellevue Manor, for always making me feel welcome no matter what state I turned up in. To Eve, it was a big help to share the end journey with you. To my friends from Sydney Summer Series, the Sydney Bush Walkers and the UTS Outdoor Adventure Club, thank you for the many adventures that got me away from the skyscrapers and helped me keep my sanity over these long years.

Finally, thank ye to Brooner. It would have been tougher without you.

## Table of Contents

Acknowledgements .....	iv
Table of Contents .....	vi
List of Tables .....	xii
List of Figures .....	xvi
List of Boxes .....	xxi
List of Abbreviations.....	xxii
Abstract.....	xxiv
<b>Chapter 1: Introduction.....</b>	<b>1</b>
1.1 Outline of the thesis .....	3
1.2 Ethics.....	5
<b>Chapter 2: Assessing the impact of vaccination programs on burden of disease: underlying complexities and statistical methods.....</b>	<b>6</b>
2.1 Chapter overview .....	7
2.2 Introduction .....	7
2.3 Estimation of burden of disease.....	9
2.4 Statistical methods for assessing disease burden over time .....	11
2.4.1 Modelling autocorrelation .....	13
2.4.2 Modelling periodicity.....	16
2.4.3 Alternative methods to model the correlation structure .....	17
2.4.4 Modelling disease burden over time using an alternative link function .....	18
2.4.5 Testing for a change in trend as a result of vaccination efforts .....	19
2.4.6 Analysing the indirect effects of vaccination programs .....	20
2.4.7 Modelling the proportion of non-specific burden of disease that is vaccine preventable .....	21

2.5 Statistical methods used in the literature for assessing disease incidence over time .....	24
2.6 Discussion.....	25
2.7 Chapter summary .....	26
<b>Chapter 3: Datasets for estimation of rotavirus and acute gastroenteritis disease burden in New South Wales, Australia using routinely collected data .....</b>	<b>28</b>
3.1 Chapter overview .....	28
3.2 Rotavirus disease.....	28
3.3 Rotavirus vaccines .....	29
3.4 Rotavirus in Australia.....	29
3.5 Rotavirus vaccination in Australia.....	31
3.6 Data sources.....	32
3.6.1 Hospital data .....	32
3.6.2 Emergency department data .....	33
3.6.3 Pathology data .....	35
3.6.4 NSW Population data.....	36
3.7 Software.....	37
3.8 Chapter summary .....	37
<b>Chapter 4: Impact of the Australian rotavirus vaccination program on acute gastroenteritis burden in NSW .....</b>	<b>38</b>
4.1 Chapter overview .....	38
4.2 Introduction.....	38
4.2.1 Chapter aims .....	48
4.3 Methods .....	48
4.3.1 Outcome variable to analyse the impact of the Australian vaccination program in NSW.....	48

4.3.2 Assessing changes in temporal trends due to vaccination efforts .....	49
4.3.3 Quantifying changes in the burden of disease due to vaccination efforts .....	53
4.4 Results .....	54
4.4.1 Crude estimates of acute gastroenteritis disease burden in NSW .....	54
4.4.2 Assessing changes in the underlying temporal trends due to vaccination efforts .....	56
4.4.3 Quantifying changes in the burden of disease.....	61
4.5 Discussion.....	64
4.6 Chapter summary .....	73
<b>Chapter 5: An investigation of statistical methods to estimate the number of acute gastroenteritis hospitalisations and emergency department presentations in NSW that are due to rotavirus.....</b>	<b>74</b>
5.1 Chapter Overview .....	74
5.2 Introduction .....	74
5.3 Methods .....	76
5.3.1 Rotavirus coded hospitalisations .....	76
5.3.2 Off-season attributable fraction method .....	76
5.3.3 Linear regression model.....	78
5.3.4 Generalised additive model .....	79
5.4 Results .....	81
5.4.1 Estimates of rotavirus hospitalisations using the A08.0 rotavirus code .....	81
5.4.2 Estimates of rotavirus-attributed hospitalisations and ED presentations using the off-season attributable fractions method .....	83
5.4.3 Estimates of rotavirus-attributed hospitalisations and ED presentations using the linear regression model.....	84



5.4.4 Estimates of rotavirus-attributed hospitalisations and ED presentations using the generalised additive models .....	85
5.5 Discussion.....	90
5.6 Chapter Summary .....	95
<b>Chapter 6: Simulations to explore methods for estimating the true rotavirus disease burden .....</b>	<b>97</b>
6.1 Chapter Overview .....	97
6.2 Introduction.....	97
6.3 Scenarios.....	101
6.3.1 Scenario One.....	105
6.3.2 Scenario Two .....	107
6.3.3 Scenario Three .....	108
6.3.4 Scenario Four.....	109
6.3.5 Scenario Five .....	110
6.3.6 Scenario Six .....	111
6.3.7 Scenario Seven.....	112
6.3.8 Scenario Eight.....	112
6.3.9 Scenario Nine.....	114
6.3.10 Scenario Ten .....	115
6.3.11 Scenario Eleven .....	116
6.3.12 Scenario Twelve.....	117
6.3.13 Scenario Thirteen .....	118
6.3.14 Scenario Fourteen .....	119
6.3.15 Scenario Fifteen .....	120
6.3.16 Scenario Sixteen.....	122
6.4 Simulations .....	122
6.5 Statistical analyses .....	124

6.5.1 Off-season attributable fraction method .....	124
6.5.2 Linear regression model with no interaction term.....	125
6.5.3 Generalised additive models with no interaction term .....	126
6.5.4 Linear regression model with an interaction with program introduction .....	128
6.5.5 Generalised additive models with an interaction with program introduction .....	130
6.5.6 Linear regression model with an interaction with season.....	132
6.5.7 Generalised additive models with an interaction with season .....	133
6.5.8 Linear regression model with an interaction with season and program introduction .....	135
6.5.9 Generalised additive models with an interaction with season and program introduction .....	136
6.5.10 Summary statistics for each scenario.....	137
6.5.11 Application of methods to the acute gastroenteritis hospitalisations and ED presentations of children under five years of age in NSW (July 2001 to June 2013) .....	138
6.6 Results .....	138
6.7 Discussion.....	163
6.8 Chapter summary .....	171
<b>Chapter 7: Thesis Summary .....</b>	<b>172</b>
7.1 Chapter summary .....	172
7.2 Summary of thesis results.....	172
7.2.1 Thesis Aim 1: To provide an outline of the broad classes of statistical methods that can be used to estimate changes in burden of disease over time that best account for the characteristics of infectious diseases.....	172

7.2.2 Thesis Aim 2: To estimate the impact of the Australian rotavirus vaccination program on acute gastroenteritis hospitalisations and emergency department presentations (non-specific disease) in children under five years of age in New South Wales, Australia, using appropriate statistical methods .....	173
7.2.3 Thesis Aim 3: To directly compare statistical methods that can separate hospitalisations and ED presentations in children under five years of age in NSW for acute gastroenteritis into those that are due to rotavirus and those that are not.....	175
7.2.4 Thesis Aim 4: To estimate the fraction of acute gastroenteritis hospitalisations and ED presentations in children under five years of age in NSW that may be due to rotavirus, and investigate whether this fraction has changed following the introduction of the rotavirus vaccine .....	176
7.3 Strengths and limitations .....	177
7.4 Recommendations .....	180
7.5 Thesis conclusion.....	181
References.....	182
Appendices.....	195
Appendix A.....	195
Appendix B.....	198
Appendix C.....	200
Appendix D.....	202
Appendix E.....	203
Appendix F .....	204
Appendix G.....	205

## List of Tables

Table 2.1	Key features of some infant immunisation programs that may affect assessment of vaccination impact .....	10
Table 2.2	Key features of non-Australian publications that assess the impact of vaccination programs on burden of disease that are discussed throughout the manuscript .....	14
Table 2.3	Key features of all unique published studies that assess the impact of a national rotavirus vaccination program introduced in 2007 on burden of disease in Australia .....	22
Table 4.1	Descriptive statistics of the weekly rates of NSW hospitalisations and ED presentations for acute gastroenteritis per 100,000 children aged less than five years by year .....	55
Table 4.2	Model output from multivariable time series regression models assessing temporal trends due to vaccination efforts in the weekly rates of NSW hospitalisations and ED presentations for acute gastroenteritis per 100,000 children aged less than five years .....	57
Table 4.3	Mean weekly observed rates and forecasted “no vaccination program” rates of acute gastroenteritis hospitalisations per 100,000 children aged less than five years by year .....	60
Table 4.4	Mean weekly observed rates and forecasted “no vaccination program” rates of acute gastroenteritis ED presentations per 100,000 children aged less than five years by year .....	61
Table 5.1	Estimated total numbers of rotavirus-attributed hospital and ED cases over the study period in children under the age of five years and as a percentage of the total number of respective acute gastroenteritis encounters using the off-season attributable fraction	

	method .....	83
Table 5.2	Output from the linear regression model .....	84
Table 5.3	Estimated total numbers of rotavirus attributed hospital and ED cases over the study period in children under the age of five years and as a percentage of the total number of respective acute gastroenteritis encounters when using linear regression models or generalised additive models .....	85
Table 5.4	Output from the generalised additive models .....	87
Table 6.1	Summary of the sixteen scenarios explored in this chapter .....	102
Table 6.2	Scenario 1 results from 1000 simulations: Ratio of rotavirus cases per positive laboratory result is constant over the study period .....	140
Table 6.3	Scenario 2 results from 1000 simulations: ratio of weekly rotavirus cases per positive laboratory result varies by program introduction (pre/post) .....	141
Table 6.4	Scenario 3 results from 1000 simulations: ratio of weekly rotavirus cases per positive laboratory result varies by season (peak and off-peak) and program introduction (pre/post) .....	142
Table 6.5	Scenario 4 results from 1000 simulations: increased non-seasonal rotavirus cases .....	143
Table 6.6	Scenario 5 results from 1000 simulations: no non-seasonal rotavirus cases .....	144
Table 6.7	Scenario 6 results from 1000 simulations: Scenario five with positive laboratory results lagged by one week .....	145
Table 6.8	Scenario 7 results from 1000 simulations: no relationship between weekly positive laboratory results and rotavirus cases (seasonal peaks and widths coincide, seasonal magnitudes vary) .....	146

Table 6.9	Scenario 8 results from 1000 simulations: no relationship between weekly positive laboratory results and rotavirus cases (seasonal peaks coincide, seasonal widths and magnitudes vary) .....	147
Table 6.10	Scenario 9 results from 1000 simulations: varying heights in peaks of rotavirus seasonality. Ratio of rotavirus cases per positive laboratory result is constant over the study period .....	148
Table 6.11	Scenario 10 results from 1000 simulations: varying heights in peaks of rotavirus seasonality. Ratio of rotavirus cases per positive laboratory result varies by year .....	149
Table 6.12	Scenario 11 results from 1000 simulations: seasonality in non-rotavirus cases with the seasonal non-rotavirus peak coinciding with the seasonal rotavirus peak .....	150
Table 6.13	Scenario 12 results from 1000 simulations: summer and winter seasonality in non-rotavirus cases .....	151
Table 6.14	Scenario 13 results from 1000 simulations: decreasing linear trend in rotavirus cases .....	152
Table 6.15	Scenario 14 results from 1000 simulations: decreasing linear trend in peaks of seasonal rotavirus cases and no effect of vaccination efforts .....	153
Table 6.16	Scenario 15 results from 1000 simulations: combination of Scenarios 10 and 11 with varying heights in peaks of non-rotavirus seasonality .....	154
Table 6.17	Scenario 16 results from 1000 simulations: combination of Scenarios 10 and 12 with varying heights in peaks of non-rotavirus seasonality .....	155
Table 6.18	Scenario 3 results from 1000 simulations with seasonal interactions incorporated into the statistical models: ratio of weekly rotavirus	

	cases per positive laboratory result varies by season (peak and off-peak) and program introduction (pre/post) .....	160
Table 6.19	Observed percentages and model coefficients for the hospital data by method .....	161
Table 6.20	Observed percentages and model coefficients for the ED data by method .....	162

## List of Figures

Figure 4.1	Hypothetical scenarios where an independent samples t-test can be used to test for an impact of the vaccination program. Panel A shows a stable disease burden over time where an immediate reduction in disease burden occurs with introduction of the vaccination program. Panel B is similar to Panel A but with additional periodicity that remains constant in the size and timing of the peaks over time .....	40
Figure 4.2	Hypothetical scenario where the rate of rotavirus hospitalisations was 150 per 100,000 population in the first week and there was a decreasing weekly trend in this rate of 0.1 per 100,000 population that was not affected by vaccination efforts .....	41
Figure 4.3	Hypothetical scenario where vaccination efforts cause the underlying trend to change (Panel A), the periodicity to change (Panel B), the periodicity to shift (Panel C), or all of the above (Panel D) .....	42
Figure 4.4	Hypothetical scenario where vaccination efforts cause the underlying trend to decline faster and the size of the seasonal peaks to halve .....	44
Figure 4.5	Hypothetical scenario from figure 4.4 that shows the change in underlying trend and seasonality due to vaccination efforts .....	45
Figure 4.6	Hypothetical scenario from figure 4.4 that forecasts a “no vaccination program” scenario into the vaccination era using pre-vaccination era data only.....	46
Figure 4.7	Sinusoidal curve demonstrating annual seasonality .....	50
Figure 4.8	Weekly rates of NSW hospitalisations and non-admitted ED	



	presentations for all cause acute gastroenteritis in children under the age of 5 years in NSW, Australia, 2001-2013 .....	54
Figure 4.9	Observed and predicted weekly rates of NSW hospitalisations from the model shown in Table 4.2 .....	58
Figure 4.10	Estimated periodicity and underlying trend in weekly rates of NSW hospitalisations from the model shown in Table 4.2 .....	58
Figure 4.11	Observed and predicted weekly rates of NSW ED presentations from the model shown in Table 4.2 .....	59
Figure 4.12	Estimated periodicity and underlying trend in weekly rates of NSW ED presentations from the model shown in Table 4.2 .....	59
Figure 4.13	Observed and forecasted weekly rates of NSW hospitalisations per 100,000 children aged under five years of age (July 2001-December 2010) .....	62
Figure 4.14	Observed and forecasted weekly rates of NSW hospitalisations per 100,000 children aged under five years of age (July 2007- July 2013) .....	62
Figure 4.15	Observed and forecasted weekly rates of NSW ED presentations per 100,000 children aged under five years of age (July 2001-December 2010) .....	63
Figure 4.16	Observed and forecasted weekly rates of NSW ED presentations per 100,000 children aged under five years of age (July 2007- July 2013) .....	63
Figure 5.1	Breakdown of the attributable fraction method showing the components of acute gastroenteritis cases and rotavirus cases over the study period .....	77
Figure 5.2	Weekly numbers of positive rotavirus laboratory results in the Pathology West Laboratory Network in children under the age of 5	

	years, 2001-2013 .....	82
Figure 5.3	Weekly numbers of hospitalisations with a primary rotavirus coded (A08.0) diagnosis in children under the age of 5 years in NSW, Australia, 2001-2013 .....	82
Figure 5.4	Estimated weekly numbers of rotavirus and non-rotavirus attributed acute gastroenteritis hospitalisations using a linear regression model for children under five years of age .....	86
Figure 5.5	Estimated weekly numbers of rotavirus and non-rotavirus attributed acute gastroenteritis ED presentations using a linear regression model for children under five years of age .....	86
Figure 5.6	Estimated weekly numbers of rotavirus and non-rotavirus attributed acute gastroenteritis hospitalisations using a generalised additive model with 6 knots per year .....	88
Figure 5.7	Estimated weekly numbers of rotavirus and non-rotavirus attributed acute gastroenteritis hospitalisations using a generalised additive model with 2 knots per year .....	88
Figure 5.8	Estimated weekly numbers of rotavirus and non-rotavirus attributed acute gastroenteritis ED presentations using a generalised additive model with 6 knots .....	89
Figure 5.9	Estimated weekly numbers of rotavirus and non-rotavirus attributed acute gastroenteritis ED presentations using a generalised additive model with 2 knots .....	89
Figure 6.1	Scenario 1: ratio of weekly rotavirus cases per positive laboratory result is constant over the study period .....	106
Figure 6.2	Scenario 2: ratio of rotavirus cases per positive laboratory result varies by program introduction (pre/post) .....	107
Figure 6.3	Scenario 3: ratio of weekly rotavirus cases per positive laboratory	

	result varies by season (peak/off-peak) and program introduction (pre/post) .....	108
Figure 6.4	Scenario 4: increase in non-seasonal rotavirus cases .....	109
Figure 6.5	Scenario 5: no non-seasonal rotavirus cases .....	110
Figure 6.6	Scenario 6: Scenario 5 with positive laboratory results lagged by one week .....	111
Figure 6.7	Scenario 7: no relationship between weekly positive laboratory results and rotavirus cases (seasonal peaks and widths coincide, seasonal magnitudes vary) .....	113
Figure 6.8	Scenario 8: no relationship between weekly positive laboratory results and rotavirus cases (seasonal peaks coincide, seasonal widths and magnitudes vary) .....	113
Figure 6.9	Scenario 9: varying heights in peaks of rotavirus seasonality. Ratio of weekly rotavirus cases per positive laboratory result is constant over the study .....	114
Figure 6.10	Scenario 10: varying heights in peaks of rotavirus seasonality. Ratio of weekly rotavirus cases per positive laboratory result varies by year .....	115
Figure 6.11	Scenario 11: seasonality in non-rotavirus cases with the seasonal non-rotavirus peak coinciding with the seasonal rotavirus peak .....	116
Figure 6.12	Scenario 12: summer and winter seasonality in non-rotavirus cases .....	117
Figure 6.13	Scenario 13: decreasing linear trend in rotavirus cases .....	118
Figure 6.14	Scenario 14: decreasing linear trend in peaks of seasonal rotavirus cases and no effect of vaccination efforts .....	119
Figure 6.15	Scenario 15: combination of Scenarios 10 and 11 with varying	

	heights in peaks of non-rotavirus seasonality .....	121
Figure 6.16	Scenario 16: combination of Scenarios 10 and 12 with varying heights in peaks of non-rotavirus seasonality .....	121
Figure 6.17	Observed numbers of positive laboratory counts versus the rotavirus cases in the same week by pre- and post-program era from simulation 15 of scenario 1 .....	158
Figure 6.18	Observed numbers of positive laboratory counts versus the rotavirus cases in the same week by pre- and post-program era from simulation 15 of scenario 2 .....	158
Figure 6.19	True numbers of positive laboratory counts versus the rotavirus cases in the same week by pre-/post-program era and season for scenario 3 .....	159
Figure 6.20	Observed numbers of positive laboratory counts versus the rotavirus cases in the same week by pre-/post-program era and season from simulation 15 of scenario 3.....	159

## List of Boxes

Box 6.1	Summary of the simulation study process carried out in this chapter	
	.....	100

## List of Abbreviations

ACT	Australian Capital Territory
AGE	Acute Gastroenteritis
AIC	Akaike's Information Criterion
APDC	Admitted Patient Data Collection
AR	Autoregressive
ARMA	Autoregressive Moving Average
ARIMA	Autoregressive Integrated Moving Average
CI	Confidence Interval
ED	Emergency Department
EDDC	Emergency Department Data Collection
GP	General Practitioner
HBeAg	Hepatitis B Envelope Antigen
HBsAg	Hepatitis B Surface Antigen
ICD-9	International Classification of Diseases Ninth Revision
ICD-10	International Classification of Diseases Tenth Revision
ICD-10-AM	International Classification of Diseases Tenth Revision Australian Modification
MA	Moving Average
NSW	New South Wales
NT	Northern Territory
PCV7	7-valent Pneumococcal Conjugate Vaccine
QQ-plot	Quantile-quantile plot
Qld	Queensland
RCT	Randomised Controlled Trial
RSV	Respiratory Syncytial Virus

RV	Rotavirus
SA	South Australia
SNOMED CT	Systemised Nomenclature of Medicine Clinical Terms
US	United States
Vic	Victoria
WA	Western Australia

## **Abstract**

Vaccination programs can provide an effective means to control infectious diseases at a population level. Evaluating the impact of these programs after implementation can be complicated by limitations of routine surveillance systems and lack of routine testing to confirm diagnosis, as well as natural fluctuations in disease rates over time. This thesis provides a structured explanation of statistical methods and how they can be used to address the epidemiological challenges in assessing changes in burden of disease as a result of vaccination programs. It explores statistical approaches to the evaluation of vaccination programs at a population level, using rotavirus vaccination in New South Wales (NSW), Australia, as a case study.

A summary of the key features that need to be considered when trying to detect any changes in the burden of infectious diseases due to vaccination is given. An assessment of the impact of the Australian rotavirus vaccination program, introduced in 2007, on hospitalisations and emergency department (ED) presentations of children aged under five years for all-cause gastroenteritis in NSW from July 2001 to June 2013 was conducted. Previously published methods that separate these hospitalisations and ED presentations into those due to rotavirus and those that are not are compared. A simulation study is used to explore these methods in controlled scenarios to determine the most appropriate method for these data.

The Australian rotavirus vaccination program had an almost immediate impact and led to a fifty and sixteen percent reduction in the rate of hospitalisations and ED presentations for acute gastroenteritis within the first 2.5 years. These declines were mostly attributed to a decline in the size of the seasonal peak. The methods to determine rotavirus cases from all-cause



gastroenteritis cases had different strengths and limitations and the derived estimates varied. No robust method was identified from the simulation study for our data. Each method that relied on using weekly counts of positive rotavirus laboratory tests to estimate rotavirus-attributable cases underestimated the true number of rotavirus cases when their assumptions held.

The evaluation of the effectiveness of vaccination programs requires the use of rigorous statistical methods to ensure the robustness and validity of findings. Appropriate statistical methods that account for temporal trends are needed to provide a detailed understanding of any changes in disease burden observed. While this thesis focused on rotavirus disease burden in NSW, Australia, many of the concepts discussed are applicable to other infectious diseases.

